

An Efficient Algorithm for Maximum-Entropy Extension of Block–Circulant Covariance Matrices

Francesca P. Carli, Augusto Ferrante, Michele Pavon, and Giorgio Picci

Abstract

This paper deals with maximum entropy completion of partially specified block–circulant matrices. Since positive definite symmetric circulants happen to be covariance matrices of stationary periodic processes, in particular of stationary *reciprocal processes*, this problem has applications in signal processing, in particular to image modeling. Maximum entropy completion is strictly related to maximum likelihood estimation subject to certain conditional independence constraints. The maximum entropy completion problem for block–circulant matrices is a nonlinear problem which has recently been solved by the authors, although leaving open the problem of an efficient computation of the solution. The main contribution of this paper is to provide an efficient algorithm for computing the solution. Simulation shows that our iterative scheme outperforms various existing approaches, especially for large dimensional problems. A necessary and sufficient condition for the existence of a positive definite circulant completion for unitary bandwidth and block–size is also provided.

I. INTRODUCTION

We consider the problem of completing a partially specified block–circulant matrix under the constraint that the completed matrix should be positive definite and block–circulant with an

Work partially supported by the Italian Ministry for Education and Research (MIUR) under PRIN grant “Identification and Adaptive Control of Industrial Systems”.

F. P. Carli, A. Ferrante and G. Picci are with the Department of Information Engineering (DEI), University of Padova, via Gradenigo 6/B, 35131 Padova, Italy. carlifra@dei.unipd.it, augusto@dei.unipd.it, picci@dei.unipd.it

M. Pavon is with the Department of Pure and Applied Mathematics, University of Padova, pavon@math.unipd.it

inverse of banded structure. As shown in [4] a block–circulant completion problem of this kind is a crucial tool for the identification of a class of reciprocal processes. These processes, [22], [25], [27], are a generalization of Markov processes which are particularly useful for modeling random signals which live in a finite region of time or of the space line (think for example of an image). In this paper we consider *stationary* reciprocal processes for which we refer the reader to [24], [13] and references therein. In particular, stationary reciprocal processes of the autoregressive type can be described by linear models involving a banded block–circulant concentration matrix¹ whose blocks are the (matrix–valued) parameter of the model.

This problem fits in the general framework of covariance extension problems introduced by A. P. Dempster [11] and studied by many authors (see [12], [20], [10], [28], [18], [1], [17], [14] and references therein). A key discovery by Dempster is that the inverse of the maximum entropy completion has zeros in the positions corresponding to the unspecified entries in the original partially given covariance matrix, a property which, from now on, will be referred to as the Dempster property. For the special case of block-Toeplitz matrices the inverse of the maximum entropy extension turns out to have a banded structure. See [12] and [17] for a general formulation of matrix extension problems in terms of banded algebras and for a thorough discussion of the so-called *band-extension* problem for block-Toeplitz matrices. It turns out that the block-Toeplitz band extension problem can be solved by factorization techniques and is essentially a linear problem. This is unfortunately no longer true when a block- circulant structure is imposed [6] to the extension.

A relevant fact is that, even when the constraint of a circulant structure is imposed, the inverse of the maximum entropy completion maintains the Dempster property. In fact, it has been first noticed in [4] for a particular data structure and then proved in complete generality in [5], that if the data are compatible with a block-circulant structure then the maximum entropy completion turns out to be automatically block-circulant. Otherwise stated, the solution of the Maximum Entropy block-Circulant Extension Problem (CMEP) and of the Dempster Maximum Entropy Extension Problem (DMEP) with data consistent with a block–circulant structure, coincide. This fact permits to simplify the formulation of the entropy optimization problem considerably. Note that this property does not hold, for example, for arbitrary missing elements in a Toeplitz structure

¹i.e. the inverse covariance matrix, also known as the precision matrix.

since in such a case the completion is not automatically Toeplitz unless the given data lie on consecutive bands centered along the main diagonal [12].

Since the solutions of the CMEP and of the DMEP with circulant-compatible data coincide, all methods available in the literature for the DMEP can, in principle, be employed to compute the solution of CMEP. In particular, it has been shown that if the graph associated with the specified entries is chordal ([19]), the solution of the DMEP can be expressed in closed form in terms of the principal minors of the sample covariance matrix; see [1], [15], [26]. The sparsity pattern associated with the given entries in our problem is not chordal. For non-chordal graphs the maximum entropy completion has to be computed iteratively. A straightforward application of standard optimization algorithms is too expensive for large size problems, and several specialized algorithms have been proposed in the literature; see [11], [28], [29], [23]. These algorithms, however, deal with the general unstructured setting of Dempster. In the present work, we propose a modified matricial gradient descent algorithm for the solution of the CMEP which naturally follows from the variational analysis in [4] and exploits in an essential way the circulant structure of our problem. This algorithm compares very favourably with the algorithms proposed in the literature for the solution of the DMEP.

A further issue addressed in the present paper is the feasibility of the CMEP. In [5], a description of the set of all positive definite completions of a partially specified block-circulant matrix has been provided, while in [4] a sufficient condition on the data for a positive definite block-circulant completion to exist has been derived. Here we provide a necessary and sufficient condition for the feasibility of the CMEP for unitary bandwidth and block-size.

Necessary and sufficient conditions for the existence of a positive definite completion have also been discussed by Grone, Johnson, Sa and Wolkowicz [20]. Although our results may seem different they are in fact not in contrast with those in [20]. See Remark 6.2 in Section VI.

The outline of the paper is as follows. In Section II we introduce some notation and state the entropy maximization problem. Section III contains a brief review of some of the most popular methods for the solution of the DMEP. In Section IV the matricial gradient descent algorithm is introduced. An experimental comparison between the proposed algorithm and the methods available in the literature is presented in Section V. Finally, in Section VI a necessary and sufficient condition for the feasibility of the CMEP is discussed. A numerical example which sheds light on the feasibility issue is presented in Section VII.

II. NOTATION AND PRELIMINARIES

All random variables in this paper, denoted by boldface characters, have zero mean and finite second order moments. It is shown in [4] that a wide-sense stationary \mathbb{R}^m -valued process \mathbf{y} is stationary reciprocal on $[0, N]$ if and only if its covariance matrix, say Σ_N , has a block-circulant symmetric structure, i.e. Σ_N is of the form

$$\Sigma_N = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_\tau^\top & \dots & \Sigma_\tau & \dots & \Sigma_1 \\ \Sigma_1 & \tilde{\Sigma}_0 & \Sigma_1^\top & \ddots & \Sigma_\tau^\top & \dots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & & \Sigma_\tau \\ \Sigma_\tau & \dots & \Sigma_1 & \Sigma_0 & \Sigma_1^\top & \dots & \ddots & \\ \vdots & \Sigma_\tau & \dots & & \Sigma_0 & & \dots & \Sigma_\tau^\top \\ \Sigma_\tau^\top & & \ddots & & & & & \vdots \\ \vdots & \ddots & & \ddots & & \ddots & \ddots & \Sigma_1^\top \\ \Sigma_1^\top & \dots & \Sigma_\tau^\top & \dots & \Sigma_\tau & & \Sigma_1 & \Sigma_0 \end{bmatrix}.$$

where $\Sigma_k = \mathbb{E} \mathbf{y}(t+k) \mathbf{y}(t)^\top$. We refer the reader to [9] for an introduction to circulants; an extension of some relevant results for the block-case can be found, for example, in [5]. Here we just recall that the class of circulants is closed under sum, product, inverse and transpose. Moreover, all circulants commute and are simultaneously diagonalized by the Fourier matrix, see (21) below, of suitable size.

Let \mathfrak{S}_N denote the vector space of real *symmetric* matrices with $N \times N$ square blocks of dimension $m \times m$. Moreover, let \mathcal{I}_b be the set of pairs of indices consistent with a banded-symmetric block-circulant structure of bandwidth n , i.e. the set of the (i, j) 's such that

$$\begin{aligned} \text{if } |i - j| \leq mn &\Rightarrow (i, j) \in \mathcal{I}_b \\ \text{if } (i, j) \in \mathcal{I}_b &\Rightarrow (j, i) \in \mathcal{I}_b \\ \text{if } (i, j) \in \mathcal{I}_b &\Rightarrow ((j + m)_{\text{mod } mN}, (i + m)_{\text{mod } mN}) \in \mathcal{I}_b. \end{aligned}$$

(an example of this structure is shown in Figure 1a). The *differential entropy* $H(p)$ of a probability density function p on \mathbb{R}^n is defined by

$$H(p) = - \int_{\mathbb{R}^n} \log(p(x)) p(x) dx. \quad (2)$$

In case of a zero-mean Gaussian distribution p with covariance matrix Σ_N , it results

$$H(p) = \frac{1}{2} \log(\det \Sigma_N) + \frac{1}{2} n (1 + \log(2\pi)). \quad (3)$$

The Maximum Entropy covariance extension problem (CMEP) for block-circulant matrices can be stated as follows.

$$\max \{ \log \det \Sigma_N \mid \Sigma_N \in \mathfrak{S}_N, \Sigma_N > 0 \} \quad (4a)$$

subject to :

$$e_i \Sigma_N e_j^\top = r_{ij}, \text{ for } (i, j) \in \mathcal{I}_b \text{ and } r_{ij} \in \mathcal{R}_b \quad (4b)$$

$$\Sigma_N \text{ is block-circulant} \quad (4c)$$

where e_k is the row-vector $e_k = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]$ and \mathcal{R}_b denotes the set of the given data consistent with a banded-symmetric block-circulant structure, i.e.

$$\mathcal{R}_b := \{ r_{ij} \in \mathbb{R} \mid (i, j) \in \mathcal{I}_b, r_{ij} = r_{ji}, r_{ij} = r_{(j+m) \bmod mN, (i+m) \bmod mN} \}.$$

We are interested in the case where the r_{ij} 's represent the entries of a partially specified block-circulant covariance matrix, say \mathbf{R}_N . If we remove constraint (4c), we get the maximum entropy covariance selection problem studied by A. P. Dempster (DMEP).

Notice that, although in Problem 4 we are maximizing the entropy functional over zero-mean *Gaussian* densities, we are *not* actually restricting ourselves to the case of Gaussian distributions. Indeed, the Gaussian distribution with (zero mean and) covariance matrix solving (4) maximizes the entropy functional (2) over the larger family of (zero mean) probability densities whose covariance matrix satisfies the boundary conditions (4b), (4c), see [4, Theorem 7.2].

The first question we need to address is *feasibility* of the CMEP. In [4], a sufficient condition for the existence of a positive definite block-circulant completion has been provided, while a characterization of the set of all positive definite completions of a partially specified block-circulant matrix has been derived in [5]. In the present paper, we provide a necessary and sufficient condition for the feasibility of the CMEP for unitary bandwidth and block-size. For the sake of continuity in exposition, we postpone this issue until Section VI.

III. ALGORITHMS FOR THE COVARIANCE SELECTION PROBLEM

Before discussing some of the algorithms proposed in the literature for the covariance selection problem, a brief digression into some basics of graph theory is needed. First of all, it is natural to

describe the pattern of the specified entries of an $Nm \times Nm$ partial symmetric matrix $M = (m_{ij})$ by an undirected graph of Nm vertices which has an edge joining vertex i and vertex j if and only if the entry m_{ij} is specified. Since the diagonal entries are all assumed to be specified, we ignore loops at the vertices. The undirected graph will be denoted by $\mathcal{G} = (V, E)$, where V is the vertex set and E the edge set which consists of unordered pairs of distinct vertices. In any undirected graph we say that 2 vertices $u, v \in V$ are *adjacent* if $(u, v) \in E$. For any vertex set $S \subseteq V$, consider the edge set $E(S) \subseteq E$ given by

$$E(S) := \{(u, v) \in E \mid u, v \in S\}$$

The graph $\mathcal{G}(S) = (S, E(S))$ is called *subgraph of \mathcal{G} induced by S* . An induced subgraph $\mathcal{G}(S)$ is *complete* if the vertices in S are pairwise adjacent in \mathcal{G} . In this case, we say that S is *complete in \mathcal{G}* .

Definition 3.1: A *clique* is a complete subgraph that is not contained within another complete subgraph.

The sequence v_0, v_1, \dots, v_k with $(v_i, v_{i+1}) \in E$, $v_i \neq v_j$ for $i \neq j$ is said a *simple path* (of length k) in \mathcal{G} . Similarly, the sequence $v_0, v_1, \dots, v_k, v_0$ is called a *simple cycle* in \mathcal{G} . A *chord* of a path (cycle) is any edge joining two nonconsecutive vertices of the path (cycle).

Definition 3.2: An undirected graph is *chordal* if every cycle of length greater than three has a chord.

Finally, we define the complementary graph of $\mathcal{G} = (V, E)$ as the graph $\tilde{\mathcal{G}}$ with vertex set V and edge set \tilde{E} with the property that $(u, v) \in \tilde{E}$ if and only if $u \neq v$ and $(u, v) \notin E$. As anticipated in the Introduction, if the graph of the specified entries is *chordal* ([19]), the maximum determinant matrix completion problem admits a closed form solution in terms of the principal minors of the sample covariance matrix (see [1], [15], [26]). However, the graph associated with a banded circulant sparsity patterns is not chordal, as it is apparent from the example of Figure 1b. Therefore we have to resort to *iterative algorithms*. For the applications we have in mind, we are dealing with vector-valued processes possibly defined on a quite large interval. A straightforward application of standard optimization algorithms is too expensive for problems of such a size, and several specialized algorithms have been proposed in the literature ([11], [28], [29], [23]) which deal with the general, unstructured setting of the DMEP. In his early work ([11]), Dempster proposed two iterative algorithms which however are very demanding from a

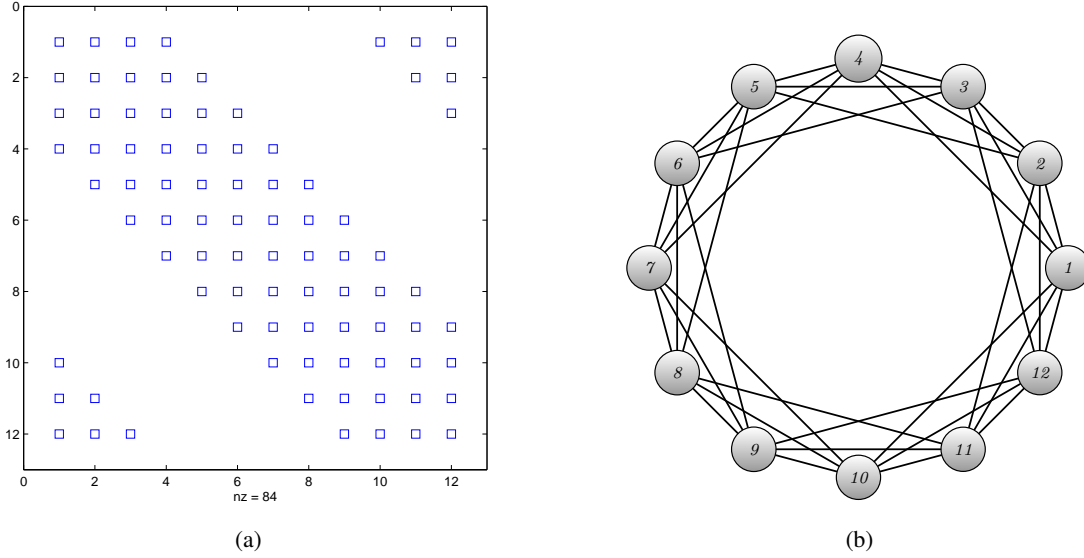


Fig. 1: Banded pattern representing the given entries for the CMEP in a block-circulant structure with $N = 12$, $n = 3$, $m = 1$ (on the left) and associated graph (on the right). The graph is not chordal since, for example, the cycle $\{1, 4, 7, 10\}$ does not have a chord.

computational point of view. Two popular methods are those proposed by T. P. Speed and H. T. Kiiveri in [28], that we now briefly discuss.

a) First algorithm: As mentioned in the Introduction, for the class of problems studied by Dempster, the inverse of the unique completion which maximizes the entropy functional has the property to be zero in the complementary positions of those fixed in Σ_N . Thus, a rather natural procedure to compute the solution of the covariance selection problem seems to be the following: iterate maintaining the elements of Σ_N in \mathcal{I}_b at the desired value (i.e. equal to the corresponding elements in the sample covariance matrix) while forcing the elements of Σ_N^{-1} in \mathcal{I}_b^c to zero. This procedure can be formally described by Algorithm 1

Algorithm 1 First algorithm (Speed and Kiiveri [28])

Compute all the cliques \tilde{c}_t in the complementary graph $\tilde{\mathcal{G}}$, say $\{\tilde{c}_t, t = 1, \dots, n_{\tilde{c}_t}\}$;

Initialize $\Sigma_N^{(0)} = \mathbf{R}_N$;

while some stopping criterion is satisfied **do**

for all the cliques \tilde{c}_t in $\tilde{\mathcal{G}}$ **do**

end for

$$\Sigma_N^{(t)} = \Sigma_N^{(t-1)} + \phi \left(\Sigma_N^{(t-1)} \right)$$

end while

where $\phi \left(\Sigma_N^{(t-1)} \right)$ is the $mN \times mN$ zero matrix which equals

$$\left\{ \text{diag} \left[\left(\left(\Sigma_N^{(t-1)} \right)^{-1} \right)_{\tilde{c}_t} \right]^{-1} \right\}^{-1} - \left[\left(\left(\Sigma_N^{(t-1)} \right)^{-1} \right)_{\tilde{c}_t} \right]^{-1}$$

in the positions corresponding to the current clique \tilde{c}_t (given a $mN \times mN$ matrix M and a set $a \subset [1, \dots, Nm]$, M_a denotes the submatrix with entries $\{m_{ij} : i, j \in a\}$). Every cycle consists of as many steps as the cliques in the complementary graph $\tilde{\mathcal{G}}$ (the graph associated to the elements in \mathcal{I}_b^c). At each step, only the elements in Σ_N corresponding to the current clique \tilde{c}_t (i.e. only a subset of the entries in \mathcal{I}_b^c) are modified in such a way to set the elements of Σ_N^{-1} in the corresponding positions to the desired zero-value. Throughout the iterations, the elements in $\Sigma_N^{(t)}$ are fixed over \mathcal{I}_b , while the elements of $\left(\Sigma_N^{(t)} \right)^{-1}$ vary over \mathcal{I}_b^c . The crucial step in the algorithm involves going from $\Sigma_N^{(t-1)}$ to $\Sigma_N^{(t)}$ and relies on the following Lemma, whose proof can be found in ([28, Lemma 2, (i)]).

Lemma 3.1 (Speed and Kiiveri): Let Q , R and B be positive definite matrices. Then for $a \subseteq V$ the matrix

$$Q^{-1} = R^{-1} + \left[\begin{array}{c|c} (B_a)^{-1} - (R_a)^{-1} & 0 \\ \hline 0 & 0 \end{array} \right] \quad (5)$$

is positive definite and satisfies

$$Q(\alpha, \beta) = B(\alpha, \beta) \quad \text{if } \alpha \in a \text{ and } \beta \in a \quad (6a)$$

$$Q^{-1}(\alpha, \beta) = R^{-1}(\alpha, \beta) \quad \text{if } \alpha \notin a \text{ and } \beta \notin a \quad (6b)$$

b) *Second algorithm:* The role of Σ_N and Σ_N^{-1} can also be swapped, yielding an alternative procedure, which is the analog of iterative proportional scaling (IPS) for contingency tables [21]. Let $\varphi\left(\Sigma_N^{(t-1)}\right)$ be the $mN \times mN$ zero matrix which equals

$$((\mathbf{R}_N)_{c_t})^{-1} - \left(\left(\Sigma_N^{(t-1)} \right)_{c_t} \right)^{-1}$$

in the positions corresponding to the current clique c_t in \mathcal{G} (the graph associated with the given entries). The second algorithm reads as follows.

Algorithm 2 Second algorithm (Speed and Kiiveri [28])

Compute all the cliques c_t in \mathcal{G} , say $\{c_t, t = 1, \dots, n_{c_t}\}$;

Initialize $\Sigma_N^{(0)} = I_{Nm}$;

while some stopping criterion is satisfied **do**

for all the cliques c_t in \mathcal{G} **do**

$$\left(\Sigma_N^{(t)} \right)^{-1} = \left(\Sigma_N^{(t-1)} \right)^{-1} + \varphi\left(\Sigma_N^{(t-1)}\right)$$

end for

end while

Every cycle consists of as many steps as the cliques in the graph of the specified entries \mathcal{G} . At each step, only the elements in Σ_N^{-1} corresponding to the current clique c_t (i.e. only a subset of the entries in \mathcal{I}_b) are modified in such a way to set the elements of Σ_N in the corresponding positions to the desired value, namely equal to the sample covariance \mathbf{R}_N (see (6a)). Through the iterations the elements in $\left(\Sigma_N^{(t)} \right)^{-1}$ are fixed over \mathcal{I}_b^c while the elements of $\Sigma_N^{(t)}$ vary over \mathcal{I}_b .

Algorithms 1 and 2 can be seen as special cases of a general procedure whose convergence is proved in [28]. An intuitive, geometric interpretation is also possible. To see this, let us denote

with \mathcal{C} ($\tilde{\mathcal{C}}$) the class of the cliques of \mathcal{G} ($\tilde{\mathcal{G}}$) and let us define the “subspaces”

$$\mathcal{P}_{L,c} = \{P > 0 \mid P_c = L_c\}$$

$$\mathcal{Q}_{M,\tilde{c}} = \{Q > 0 \mid (Q^{-1})_{\tilde{c}} \text{ agrees with } M_{\tilde{c}} \text{ except on the diagonal}\}$$

$$\mathcal{P}_{L,\mathcal{C}} = \bigcap_{c \in \mathcal{C}} \mathcal{P}_{L,c}$$

$$\mathcal{Q}_{M,\tilde{\mathcal{C}}} = \bigcap_{\tilde{c} \in \tilde{\mathcal{C}}} \mathcal{Q}_{M,\tilde{c}}$$

where, for every $c \subseteq [1, \dots, mN]$, P_c is the submatrix $P_c := \{P(i, j) \mid i \in c, j \in c\}$. In the limit, $\Sigma_N \in \mathcal{P}_{\mathbf{R}_N, \mathcal{C}} \cap \mathcal{Q}_{I_{mN}, \tilde{\mathcal{C}}}$. The first algorithm begins with $\Sigma_N^{(0)} \in \mathcal{P}_{\mathbf{R}_N, \mathcal{C}}$ and cycles through $\tilde{c} \in \tilde{\mathcal{C}}$ \mathcal{I} -projecting the current estimate of Σ_N onto $\mathcal{P}_{\mathbf{R}_N, \mathcal{C}} \cap \mathcal{Q}_{I_{mN}, \tilde{c}}$ while the second begins with $\Sigma_N^{(0)} \in \mathcal{Q}_{I_{mN}, \tilde{\mathcal{C}}}$ and cycles through $c \in \mathcal{C}$ \mathcal{I} -projecting ([7]) the current estimate onto $\mathcal{P}_{\mathbf{R}_N, c} \cap \mathcal{Q}_{I_{mN}, \tilde{\mathcal{C}}}$. (The fact that the operation performed at each step is an \mathcal{I} -projection follows from [28, Lemma 1 part (ii)]). It follows that, in a suitable sense, both algorithms are an application of the von Neumann’s alternating projections theorem for computing the projection onto the intersection of nonorthogonal subspaces.

c) Comparison of the two algorithms: The choice of which algorithm is to be preferred depends on the application and is very much dependent on the number and size of the cliques in \mathcal{G} and $\tilde{\mathcal{G}}$. In our setting, the complexity of the graph associated with the given entries depends on the bandwidth n . In particular, for bandwidth not too large with respect to the completion size (which is the case we are interested in) the complexity of the graph associated with the given data \mathcal{G} is far lower than the complexity of its complementary (which, for small n , is almost complete), see Figure 3. The execution time of the two algorithms has been compared for a completion size $N = 30$ and a bandwidth n varying between 2 and 8. The results are shown in Figure 2 and Table I. It turns out that for n small the second algorithm (which, from now on, will be referred to as IPS) runs faster than the first, and thus has to be preferred.

d) Covariance selection via chordal embedding: In [8], Dahl, Vanderberghe and Roychowdhury propose a new technique to improve the efficiency of the Newton’s method for the covariance selection problem based on chordal embedding: the given sparsity pattern is embedded in a chordal one for which they provide efficient techniques for computing the gradient and the Hessian. The complexity of the method is dominated by the cost of forming and solving a system of linear equations in which the number of unknowns depends on the number of nonzero entries

added in the chordal embedding. For a circulant sparsity pattern, it is easy to check that the number of nonzero elements added in the chordal embedding is quite large. Hence, their method does not seem to be effective for our problem. A preconditioned conjugate gradient method has also been proposed in ([8]), but the comparison is not carried out in the present paper.

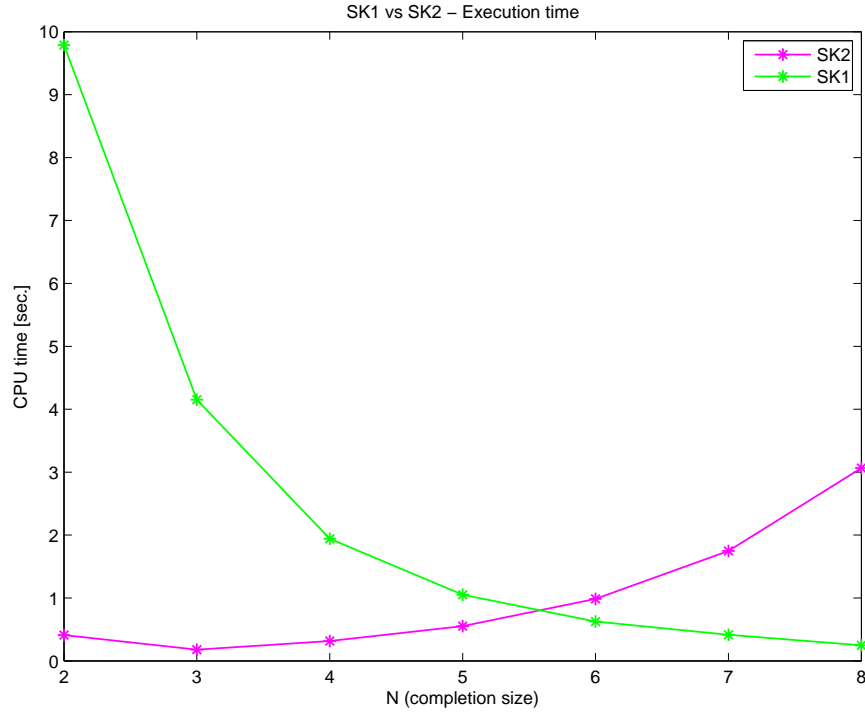


Fig. 2: Comparison between the execution time of the first and second algorithm for $N = 30$, $m = 1$, $n = \{1, \dots, 8\}$.

IV. A MATRICIAL GRADIENT DESCENT ALGORITHM

Let \mathbf{U}_N denote the block-circulant shift matrix with $N \times N$ blocks,

$$\mathbf{U}_N = \begin{bmatrix} 0 & I_m & 0 & \dots & 0 \\ 0 & 0 & I_m & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_m \\ I_m & 0 & 0 & \dots & 0 \end{bmatrix},$$

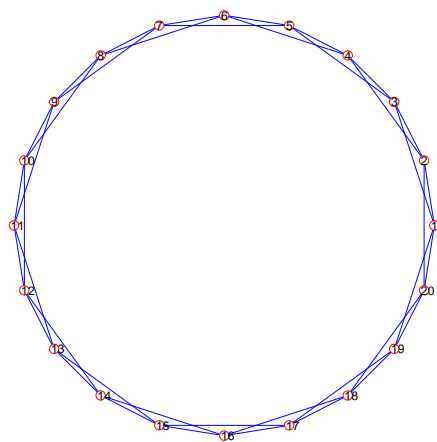
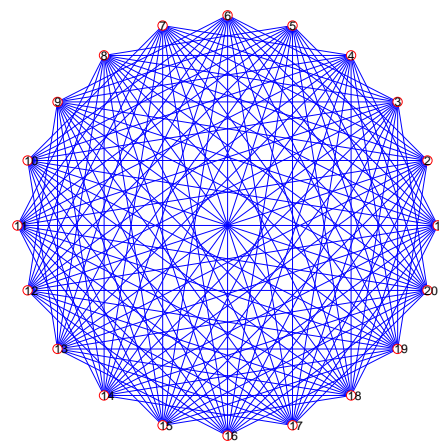
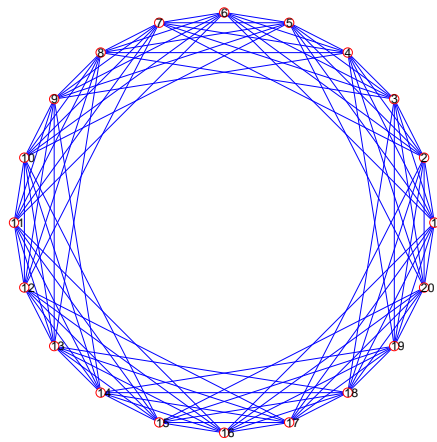
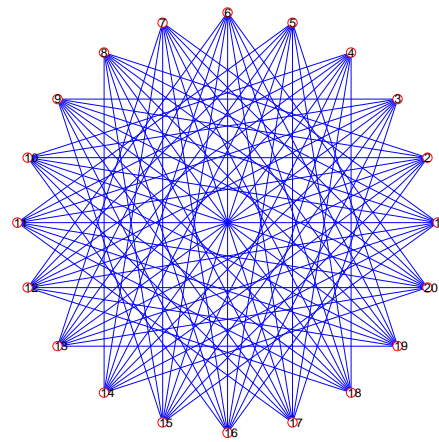
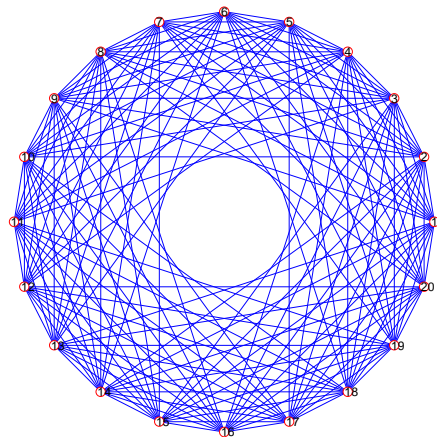
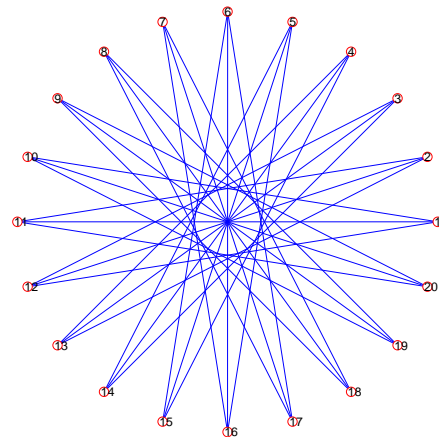
(a) \mathcal{G} for $n = 2$ (b) $\tilde{\mathcal{G}}$ for $n = 2$ (c) \mathcal{G} for $n = 5$ (d) $\tilde{\mathcal{G}}$ for $n = 5$ (e) \mathcal{G} for $n = 8$ (f) $\tilde{\mathcal{G}}$ for $n = 8$

Fig. 3: Graph \mathcal{G} associated with the given data (on the right) and its complementary $\tilde{\mathcal{G}}$ (on the left) for $N = 20$ and bandwidth $n = 2, 5, 8$.

n	First algorithm		Second algorithm	
	# of cl. (max. cl. size)	CPU time [s]	# of cl. (max. cl. size)	CPU time [s]
2	4608(10)	9.7877	30(3)	0.4109
3	2406(7)	4.1515	30(4)	0.1783
4	1241(6)	1.9419	30(5)	0.3153
5	706(5)	1.0525	30(6)	0.5535
6	445(4)	0.6258	30(7)	0.9854
7	295(3)	0.4145	30(8)	1.7477
8	175(3)	0.2480	30(9)	3.0665

TABLE I: Execution time of the first and second algorithm for $N = 30$, $m = 1$, bandwidth $n = \{2, \dots, 8\}$.

let $\mathbf{T}_n \in \mathfrak{S}_{n+1}$ denote the Toeplitz matrix of *boundary data*

$$\mathbf{T}_n = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \dots & \Sigma_n^\top \\ \Sigma_1 & \Sigma_0 & \Sigma_1^\top & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \Sigma_1^\top \\ \Sigma_n & \dots & \dots & \Sigma_1 & \Sigma_0 \end{bmatrix} \quad (7)$$

and let E_n be the $N \times (n+1)$ block matrix

$$E_n = \begin{bmatrix} I_m & 0 & \dots & 0 \\ 0 & I_m & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & I_m \\ 0 & \dots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & \dots & 0 \end{bmatrix}.$$

so that, specializing the formulation in (4), the maximum entropy band extension problem for block-circulant matrices can be stated as

$$\max \{ \log \det \Sigma_N \mid \Sigma_N \in \mathfrak{S}_N, \Sigma_N > 0 \} \quad (8a)$$

subject to :

$$E_n^\top \Sigma_N E_n = \mathbf{T}_n, \quad (8b)$$

$$\mathbf{U}_N^\top \Sigma_N \mathbf{U}_N = \Sigma_N. \quad (8c)$$

where we have exploited the characterization of block-circulant matrices as invariants under the similarity, namely $\mathbf{C}_N = \mathbf{U}_N^\top \mathbf{C}_N \mathbf{U}_N$.

Problem (8) is a *convex optimization problem* since we are minimizing a strictly convex function on the intersection of a convex cone (minus the zero matrix) with a linear manifold. We shall briefly recall the approach to this problem by duality theory following [4]. To this aim, consider the linear map

$$\begin{aligned} A : \quad \mathfrak{S}_{n+1} \times \mathfrak{S}_N &\rightarrow \mathfrak{S}_N \\ (\Lambda, \Theta) &\mapsto E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta \end{aligned} \quad (9)$$

and define the set

$$\begin{aligned} \mathcal{L}_+ := \{ (\Lambda, \Theta) \in (\mathfrak{S}_{n+1} \times \mathfrak{S}_N) \mid (\Lambda, \Theta) \in (\ker(A))^\perp, \\ (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) > 0 \}. \end{aligned} \quad (10)$$

\mathcal{L}_+ is an open, convex subset of $(\ker(A))^\perp$. Letting $\langle A, B \rangle := \text{tr} A B^\top$, the *Lagrangian function* of this problem can be written

$$\begin{aligned} L(\Sigma_N, \Lambda, \Theta) &:= -\text{tr} \log \Sigma_N + \langle \Lambda, (E_n^\top \Sigma_N E_n - \mathbf{T}_n) \rangle + \langle \Theta, (\mathbf{U}_N^\top \Sigma_N \mathbf{U}_N - \Sigma_N) \rangle \\ &= -\text{tr} \log \Sigma_N + \text{tr} (E_n \Lambda E_n^\top \Sigma_N) - \text{tr} (\Lambda \mathbf{T}_n) + \text{tr} (\mathbf{U}_N \Theta \mathbf{U}_N^\top \Sigma_N) - \text{tr} (\Theta \Sigma_N) \end{aligned}$$

and its first variation (at Σ_N in direction $\delta \Sigma_N \in \mathfrak{S}_N$) is

$$\begin{aligned} \delta L(\Sigma_N, \Lambda, \Theta; \delta \Sigma_N) &= -\text{tr} (\Sigma_N^{-1} \delta \Sigma_N) + \text{tr} (E_n \Lambda E_n^\top \delta \Sigma_N) \\ &\quad + \text{tr} ((\mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \delta \Sigma_N). \end{aligned}$$

Thus $\delta L(\Sigma_N, \Lambda, \Theta; \delta \Sigma_N) = 0$, $\forall \delta \Sigma_N \in \mathfrak{S}_N$ if and only if

$$\Sigma_N^{-1} = E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta.$$

It follows that, for each fixed pair $(\Lambda, \Theta) \in \mathcal{L}_+$, the unique Σ_N^o minimizing the Lagrangian over $\mathfrak{S}_{N,+} := \{\Sigma_N \in \mathfrak{S}_N, \Sigma_N > 0\}$ is

$$\Sigma_N^o = (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1}. \quad (11)$$

Moreover, computing the Lagrangian at $\Sigma_N = \Sigma_N^o$ results in

$$\begin{aligned} L(\Sigma_N^o, \Lambda, \Theta) &= -\text{tr} \log \left((E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1} \right) \\ &\quad + \text{tr} \left[(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \right. \\ &\quad \left. (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta)^{-1} \right] - \text{tr}(\Lambda \mathbf{T}_n) \\ &= \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta) \\ &\quad + \text{tr} I_{mN} - \text{tr}(\Lambda \mathbf{T}_n). \end{aligned}$$

This is a strictly concave function on \mathcal{L}_+ whose maximization is the *dual problem* of (CMEP).

We can equivalently consider the convex problem

$$\min \{J(\Lambda, \Theta), (\Lambda, \Theta) \in \mathcal{L}_+\}, \quad (12)$$

where J is given by

$$J(\Lambda, \Theta) = \text{tr}(\Lambda \mathbf{T}_n) - \text{tr} \log (E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta). \quad (13)$$

It can be shown ([4, Theorem 6.1]) that the function J admits a unique minimum point in $(\bar{\Lambda}, \bar{\Theta})$ in \mathcal{L}_+ .

In this section we propose a *modified gradient descent algorithm with backtracking line search* (see, e.g., [2, Ch. 9]) for the numerical solution of the dual problem (12). This task requires some care because we are working in a *matricial space*. Let $\Pi_{\mathfrak{C}_N}$ denote the orthogonal projection onto the linear subspace of symmetric, block-circulant matrices \mathfrak{C}_N . Before proceeding, we need two preliminary lemmas. The first one, gives a close form expression for the projection of $E_n \Lambda E_n^\top$ onto \mathfrak{C}_N , while the second provides a useful characterization of matrices belonging to the orthogonal complement of \mathfrak{C}_N . We refer the reader to [4] for the proof of these statements.

Lemma 4.1: *Let $\Lambda \in \mathfrak{S}_{n+1}$ be the matrix*

$$\Lambda = \begin{bmatrix} \Lambda_{00} & \Lambda_{01} & \dots & \Lambda_{0n} \\ \Lambda_{01}^\top & \Lambda_{11} & \dots & \Lambda_{1n} \\ \vdots & & \ddots & \vdots \\ \Lambda_{0n}^\top & \Lambda_{1n}^\top & \dots & \Lambda_{nn} \end{bmatrix}.$$

The orthogonal projection of $E_n \Lambda E_n^\top$ onto \mathfrak{C}_N , say Π_Λ , is given by

$$\Pi_\Lambda := \Pi_{\mathfrak{C}_N} (E_n \bar{\Lambda} E_n^\top) = \begin{bmatrix} \Pi_0 & \Pi_1^\top & \Pi_2^\top & \dots & \Pi_1 \\ \Pi_1 & \Pi_0 & \Pi_1^\top & \dots & \Pi_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Pi_2^\top & \dots & \Pi_1 & \Pi_0 & \Pi_1^\top \\ \Pi_1^\top & \Pi_2^\top & \dots & \Pi_1 & \Pi_0 \end{bmatrix}$$

with

$$\Pi_0 = \frac{1}{N} (\Lambda_{00} + \Lambda_{11} + \dots + \Lambda_{nn}), \quad (14a)$$

$$\Pi_1 = \frac{1}{N} (\Lambda_{01} + \Lambda_{12} + \dots + \Lambda_{n-1,n})^\top, \quad (14b)$$

$$\vdots \quad (14c)$$

$$\Pi_n = \frac{1}{N} \Lambda_{0n}^\top, \quad (14d)$$

while $\Pi_i = 0$, for all i in the interval $n+1 \leq i \leq N-n-1$.

Lemma 4.2: Let $M \in \mathfrak{S}_N$. Then $M \in (\mathfrak{C}_N)^\perp$ if and only if it can be expressed as

$$M = \mathbf{U}_N N \mathbf{U}_N^\top - N \quad (15)$$

for some $N \in \mathfrak{S}_N$.

Consider now the functional

$$\bar{J}(\Lambda) := \text{tr}(\Lambda T_n) - \text{tr} \log \{ \Pi_{\mathfrak{C}_N} (E_n \Lambda E_n^\top) \} \quad (16)$$

whose gradient $\nabla_\Lambda \bar{J}(\Lambda)$ is given by

$$\nabla_\Lambda \bar{J}(\Lambda) = -E_n^\top [\Pi_{\mathfrak{C}_N} (E_n \Lambda E_n^\top)]^{-1} E_n + \mathbf{T}_n.$$

The proposed algorithm is as follows.

Theorem 4.1: Algorithm 3 is a gradient descent algorithm restricted to the subspace

$$\{(\Lambda, \Theta) \mid \Pi_{\mathfrak{C}_N^\perp} (E_n \Lambda E_n^\top) = -(U_N \Theta U_N^\top - \Theta)\}. \quad (17)$$

Proof: Let $(\bar{\Lambda}, \bar{\Theta})$ be the unique minimum point of the functional J on \mathcal{L}_+ . We know that $(\bar{\Lambda}, \bar{\Theta})$ are such that $\Sigma^\circ = E_n \bar{\Lambda} E_n^\top + U_N \bar{\Theta} U_N^\top - \bar{\Theta}$ is circulant. Thus, one can think of restricting the search for the solution of the optimization problem to the set

$$\{(\Lambda, \Theta) \mid (E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta) \text{ is circulant}\}$$

Algorithm 3 Matricial gradient descent algorithm

Given a starting point $\Lambda \in \text{dom } \bar{J}$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$

while $\|\nabla_{\Lambda} \bar{J}(\Lambda)\|_2 > \eta$ **do**

$\Delta\Lambda := -\nabla_{\Lambda} \bar{J}(\Lambda)$

while $\bar{J}(\Lambda + t\Delta\Lambda) > \bar{J}(\Lambda) + \alpha t \text{tr} \{ \nabla \bar{J}(\Lambda)^\top \Delta\Lambda \}$ **do**

$t := \beta t$

end while

$\Lambda := \Lambda + t\Delta\Lambda$

end while

which can be represented as

$$\left\{ (\Lambda, \Theta) \mid \Pi_{\mathfrak{C}_N^\perp} (E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta) = 0 \right\}.$$

Since $(U_N \Theta U_N^\top - \Theta) \in \mathfrak{C}_N^\perp$ (see Lemma 4.2), this expression can be rewritten as

$$\left\{ (\Lambda, \Theta) \mid \Pi_{\mathfrak{C}_N^\perp} (E_n \Lambda E_n^\top) = -(U_N \Theta U_N^\top - \Theta) \right\}.$$

If we compute the dual function J on the set (17), we obtain

$$\begin{aligned} J(\Lambda, \Theta) & \mid \left\{ (\Lambda, \Theta) \mid \Pi_{\mathfrak{C}_N^\perp} (E_n \Lambda E_n^\top) = -(U_N \Theta U_N^\top - \Theta) \right\} \\ &= \text{tr}(\Lambda T_n) - \text{tr} \log (E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta^\top) \\ &= \text{tr}(\Lambda T_n) - \text{tr} \log \left(E_n \Lambda E_n^\top - \Pi_{\mathfrak{C}_N^\perp} (E_n \Lambda E_n^\top) \right) \\ &= \text{tr}(\Lambda T_n) - \text{tr} \log (\Pi_{\mathfrak{C}_N} (E_n \Lambda E_n^\top)) \end{aligned} \tag{18}$$

which is the modified functional defined above. This shows that the proposed algorithm is nothing but a gradient descent algorithm in which the search of the minimum point has been restricted to the subspace where the optimal solution is known to lie, i.e. subspace (17). ■

A. Initialization

The following proposition provides a good starting point for the iterative procedure of Algorithm 3.

Proposition 4.1: *Let the block-Toeplitz matrix made of the first $n+1$, $m \times m$ covariance lags $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$*

$$\mathbf{T}_n := \text{Toepl}(\Sigma_0, \Sigma_1, \dots, \Sigma_n),$$

be positive definite and let $\{\hat{\Sigma}_k, k = 0, 1, 2, \dots\}$ with

$$\hat{\Sigma}_k = \Sigma_k, \quad k = 0, 1, 2, \dots, n$$

be the maximum entropy (positive) extension of $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$. Then, for N large enough the block-circulant matrix $\Sigma_N^{(c)}$ given by

$$\begin{aligned} &\text{Toepl}\left(\hat{\Sigma}_0, \hat{\Sigma}_1^\top, \dots, \hat{\Sigma}_n^\top, \hat{\Sigma}_{n+1}^\top, \dots, \hat{\Sigma}_{\frac{N}{2}-1}^\top, \hat{\Sigma}_{\frac{N}{2}}^\top + \hat{\Sigma}_{\frac{N}{2}}, \hat{\Sigma}_{\frac{N}{2}-1}, \dots, \hat{\Sigma}_{n+1}, \hat{\Sigma}_n, \dots, \hat{\Sigma}_1\right), \quad \text{for } N \text{ even} \\ &\text{Toepl}\left(\hat{\Sigma}_0, \hat{\Sigma}_1^\top, \dots, \hat{\Sigma}_n^\top, \hat{\Sigma}_{n+1}^\top, \dots, \hat{\Sigma}_{\frac{N-1}{2}}^\top, \hat{\Sigma}_{\frac{N-1}{2}}, \dots, \hat{\Sigma}_{n+1}, \hat{\Sigma}_n, \dots, \hat{\Sigma}_1\right), \quad \text{for } N \text{ odd} \end{aligned} \quad (19)$$

is a covariance matrix which for $N \rightarrow \infty$ is arbitrarily close to the $mN \times mN$ maximum entropy block-circulant extension of \mathbf{T}_n .

Proof: That $\Sigma_N^{(c)}$ is a valid covariance matrix for N large enough follows from [4, Theorem 5.1]. To show that $\Sigma_N^{(c)}$ given by (19) tends to the maximum entropy block-circulant completion of the given data, it suffices to show that its inverse tends to be banded block-circulant for $N \rightarrow \infty$ (i.e. its off-diagonal blocks tends uniformly to zero). To this purpose, recall that $\Sigma_N^{(c)}$ can be block-diagonalized as

$$\Sigma_N^{(c)} = \mathbf{V} \Psi_N \mathbf{V}^* \quad (20)$$

where \mathbf{V} is the Fourier block-matrix whose k, l -th block is

$$V_{kl} := 1/\sqrt{N} \exp[-j2\pi(k-1)(l-1)/N] I_m \quad (21)$$

and Ψ_N is the block-diagonal matrix

$$\Psi_N := \text{diag}(\Psi_0, \Psi_1, \dots, \Psi_{N-1}), \quad (22)$$

whose diagonal blocks Ψ_ℓ , are the coefficients of the finite Fourier transform of the first block row of $\Sigma_N^{(c)}$

$$\Psi_\ell = \hat{\Sigma}_0 + e^{j\vartheta_\ell} \hat{\Sigma}_1^\top + (e^{j\vartheta_\ell})^2 \hat{\Sigma}_2^\top + \dots + (e^{j\vartheta_\ell})^{N-2} \hat{\Sigma}_2 + (e^{j\vartheta_\ell})^{N-1} \hat{\Sigma}_1, \quad (23)$$

with $\vartheta_\ell := -2\pi\ell/N$. Thus in particular

$$\left(\Sigma_N^{(c)}\right)^{-1} = \mathbf{V} \Psi_N^{-1} \mathbf{V}^*$$

where

$$\Psi_N^{-1} := \text{diag} \left(\Psi_0^{-1}, \Psi_1^{-1}, \dots, \Psi_{N-1}^{-1} \right).$$

Now, let $\Phi(z) := \hat{\Sigma}_0 + \sum_{i=1}^{\infty} \hat{\Sigma}_i z^{-i} + \left(\sum_{i=1}^{\infty} \hat{\Sigma}_i z^{-i} \right)^*$ be the spectral density matrix of the maximum entropy completion of \mathbf{T}_n . It is well-known [30] that $\Phi(z)$ can be expressed in factored form as

$$\Phi(z) = [L_n(z^{-1})]^{-1} \Lambda_n [L_n(z^{-1})]^{-*} \quad (24)$$

where $L_n(z^{-1})$ is the n -th Levinson–Whittle matrix polynomial associated with the block-Toeplitz matrix \mathbf{T}_n

$$L(z^{-1}) = \sum_{k=0}^n A_n(k) z^{-k} \quad (25)$$

with the $A_n(k)$'s and $\Lambda_n = \Lambda_n^\top > 0$ being the solutions of the Yule-Walker type equation

$$\begin{bmatrix} A_n(0) & A_n(1) & \dots & A_n(n) \end{bmatrix} \mathbf{T}_n^\top = \begin{bmatrix} \Lambda_n & 0 & \dots & 0 \end{bmatrix}. \quad (26)$$

Note that $\Phi(z)^{-1} = L_n(z^{-1})^* \Lambda_n^{-1} L_n(z)$ is a Laurent polynomial, that can be written as

$$\Phi(z)^{-1} = M_0 + (M_1 z + M_2 z^2 + \dots + M_n z^n) + (M_1 z + M_2 z^2 + \dots + M_n z^n)^*$$

Moreover, since $e^{-j\vartheta_\ell N} = 1, \forall \ell$, the Ψ_ℓ 's can be written as ²

$$\Psi_\ell = \hat{\Sigma}_0 + e^{j\vartheta_\ell} \hat{\Sigma}_1^\top + \dots + (e^{j\vartheta_\ell})^h \hat{\Sigma}_h^\top + e^{-j\vartheta_\ell} \hat{\Sigma}_1 + \dots + (e^{-j\vartheta_\ell})^h \hat{\Sigma}_h \quad (27)$$

where

$$h := \begin{cases} \frac{N-1}{2}, & N \text{ odd} \\ N/2, & N \text{ even} \end{cases}$$

Hence

$$\Psi_\ell = \Phi(e^{j\vartheta_\ell}) - [\Delta\Phi_N(e^{j\vartheta_\ell}) + \Delta\Phi_N^*(e^{j\vartheta_\ell})] \quad (28)$$

where

$$\Delta\Phi_N(z) := \sum_{i=h+1}^{\infty} \hat{\Sigma}_i z^{-i}.$$

Since the causal part of $\Phi(z)$ is a rational function with poles inside the unit circle,

$$\sup_{l=0, \dots, N-1} \|\Delta\Phi_N(e^{j\vartheta_\ell}) + \Delta\Phi_N^*(e^{j\vartheta_\ell})\| \rightarrow 0$$

²Notice that for N even $e^{j\vartheta_\ell h} = e^{-j\vartheta_\ell h} = -1$, so that $(e^{j\vartheta_\ell})^h \hat{\Sigma}_h^\top + (e^{-j\vartheta_\ell})^h \hat{\Sigma}_h = -(\hat{\Sigma}_h + \hat{\Sigma}_h^\top)$.

exponentially fast for $N \rightarrow \infty$. It follows that for $N \rightarrow \infty$

$$\Psi_\ell^{-1} \rightarrow (\Phi(e^{j\theta_\ell}))^{-1} = M_0 + M_1 e^{j\theta_\ell} + \dots + M_n (e^{j\theta_\ell})^n + M_1^\top (e^{j\theta_\ell})^{N-1} + \dots + M_n^\top (e^{j\theta_\ell})^{N-n}, \quad (29)$$

for $\ell = 0, 1, \dots, N-1$, i.e. Ψ_ℓ^{-1} tends to the finite Fourier transform of a sequence of the form

$$M_0, M_1^\top, M_2^\top, \dots, M_n^\top, 0, \dots, 0, M_n, \dots, M_1,$$

i.e. $(\Sigma_N^{(c)})^{-1}$ tends to be banded block-circulant, as claimed. \blacksquare

An alternative way to compute the maximum entropy completion of a partially specified block-*Toeplitz* matrix is via the formula [16],

$$\Phi(z) = \left(G^*(z) \mathbf{T}_n^{-1} \tilde{B} \left(\tilde{B}^* \mathbf{T}_n^{-1} \tilde{B} \right)^{-1} \tilde{B}^* \mathbf{T}_n^{-1} G(z) \right)^{-1} \quad (30)$$

where

$$G(z) = (zI - \tilde{A})^{-1} \tilde{B} \quad (31)$$

with

$$\tilde{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & & & & I \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix}. \quad (32)$$

It follows that the spectral factor $W(z) := [L_n(z^{-1})]^{-1} \Lambda_n^{\frac{1}{2}}$ has a realization

$$W(z) = C(zI - A)^{-1} B + D$$

with $D = \Lambda_n^{\frac{1}{2}}$, $C = -[A_n(n) \ A_n(n-1) \ \dots \ A_n(1)]$ and

$$A = \begin{bmatrix} 0 & I_m & 0 & 0 & \dots & 0 \\ 0 & 0 & I_m & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & I_m \\ -A_n(n) & -A_n(n-1) & \dots & \dots & \dots & -A_n(1) \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \Lambda_n^{\frac{1}{2}} \end{bmatrix},$$

The positive real part of the maximum entropy spectrum is given by

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}^\top + \frac{1}{2}\Sigma_0 \quad (33)$$

where $\bar{C}^\top = APC^\top + BD^\top$, with $P = APA^\top + BB^\top$ and the maximum entropy covariance extension results

$$\hat{\Sigma}_k = CA^{k-1}\bar{C}^\top, \quad k > n.$$

With this extension at hand, we can compute an approximation for the maximum entropy block-circulant extension as suggested by Proposition 4.1. A good starting point for our gradient descent algorithm can then be obtained from (14) assuming for Λ a Toeplitz structure.

V. NUMERICAL EXPERIMENTS

The matricial gradient descent algorithm has been implemented in Matlab. The results are shown in Figures 4 and 5 along with Tables II and III. The implementation exploits the block-circulant symmetric structure (recall, in particular, that for block-circulants the inverse can be computed efficiently by means of a Fourier transform). At each iteration, the algorithm requires the inversion of $\lceil \frac{N+1}{2} \rceil$ matrices of order m . It follows that the execution time increases as the completion size N and the block size m increase (see Figure 4 and Table II). Finally, it also increases, although to a lesser amount, for larger bandwidth n (see Figure 5 and Table III). Table IV presents a comparison between the execution times for different starting point Λ_0 . The gradient descent algorithm has been initialized to the normalized identity and by the procedure of Section IV-A. The proposed initialization acts effectively to reduce the number of iterations (and thus the computational time) to reach the minimum.

Finally, the gradient descent algorithm (GD) has been compared to the iterative proportional scaling procedure (IPS) by Speed and Kiiveri. Both algorithms are implemented in Matlab. The Bron-Kerbosch algorithm [3] has been employed for finding the cliques in the graph for IPS. The execution times for different completion size N and block size m are plotted in Figures 7 and 8 along with Tables V and VI. It can be seen that the gradient descent algorithm runs faster than the iterative proportional scaling and the gap between the two increases as N increases. Moreover, the gap becomes much more evident as m grows, making the gradient descent algorithm more attractive for applications where the process under observation is vector-valued ($m > 1$).

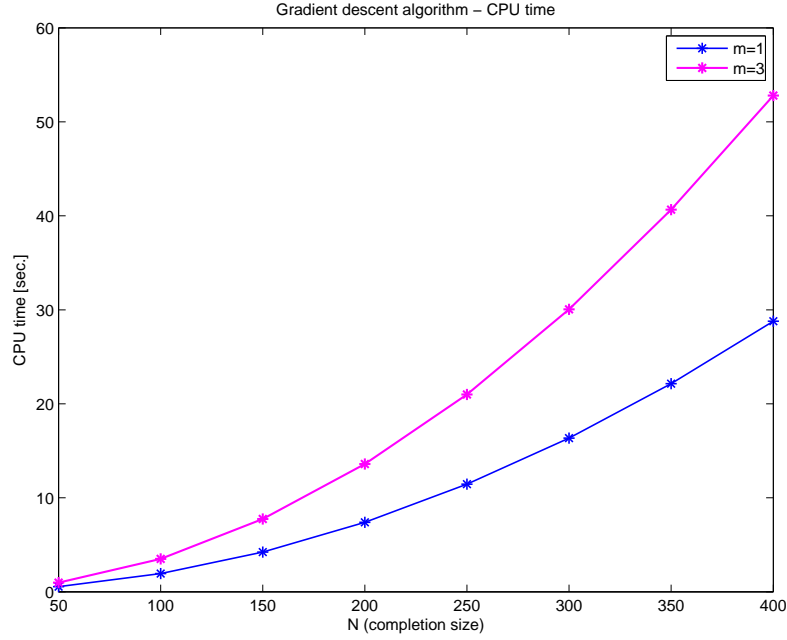


Fig. 4: Matricial gradient descent algorithm: CPU time [sec.] for bandwidth $n = 1$, $m = \{1, 3\}$, and completion size N varying from 50 to 400.

	$m = 1$	$m = 3$
50	0.5535	0.9749
100	1.9376	3.4989
150	4.2258	7.7427
200	7.3857	13.5903
250	11.4440	20.9953
300	16.3449	30.0519
350	22.1412	40.6536
400	28.7854	52.7949

TABLE II: Matricial gradient descent algorithm: CPU time [sec.] plotted in Figure 4 for bandwidth $n = 1$, $m = \{1, 3\}$, and completion size N varying from 50 to 400.

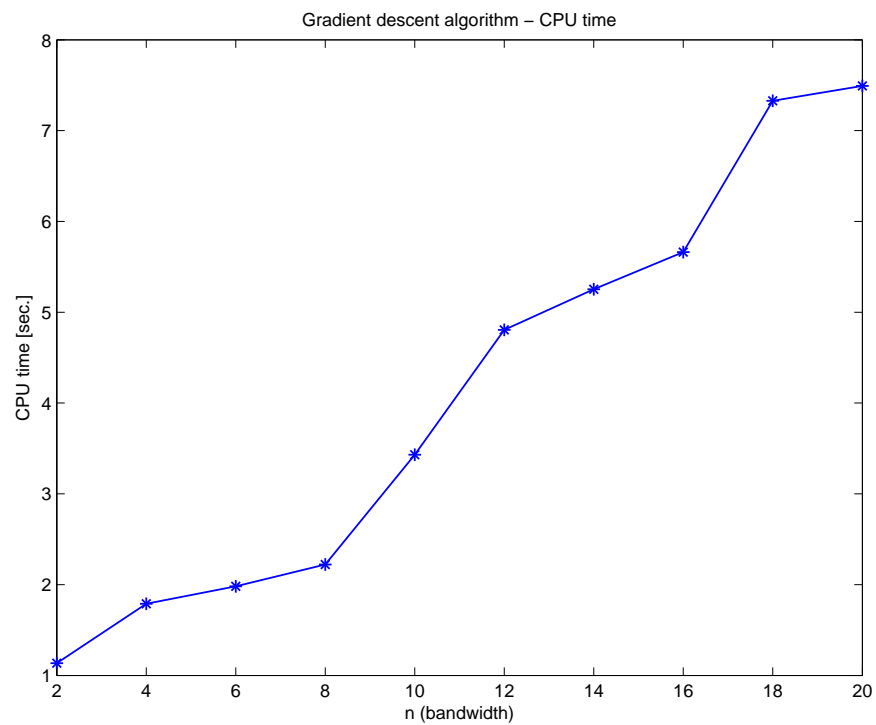


Fig. 5: Matricial gradient descent algorithm: CPU time [in sec.] for $N = 50$, $m = 1$, n varying between 2 and 20.

n	CPU time [sec.]
2	1.1351
4	1.7895
6	1.9818
8	2.2215
10	3.4312
12	4.8058
14	5.2528
16	5.6626
18	7.3284
20	7.4922

TABLE III: Matricial gradient descent algorithm: CPU time [in sec.] plotted in Figure 5 for $N = 50$, $m = 1$, n varying between 2 and 20.

N	m	GDI		GDT	
		# of itz.	CPU time	# of itz.	CPU time
10	5	99	0.1455	61	0.0767
20	5	212	0.4143	65	0.1270
30	5	322	0.8355	97	0.2504
40	5	432	1.4233	130	0.4285
50	5	541	2.1937	163	0.6603

TABLE IV: CPU time [in sec.] for the matricial gradient descent algorithm with different initializations (identity on the left and as in Section IV-A on the right). The reported times have been computed for $N = [10, 20, 30, 40, 50]$, $m = 5$ and bandwidth $n = 3$.

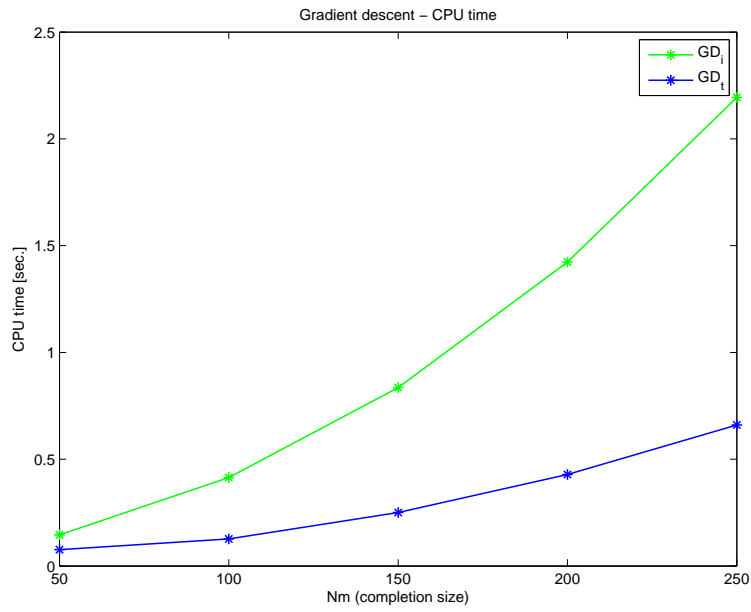


Fig. 6: CPU time [in sec.] for the matricial gradient descent algorithm with different initializations (identity in green and as in Section IV-A in blue). The reported times have been computed for $N = [10, 20, 30, 40, 50]$, $m = 5$ and bandwidth $n = 3$.

N	m	IPS	GD
10	5	4.7048	0.0767
20	5	16.4981	0.1270
30	5	29.2779	0.2504
40	5	43.8072	0.4285
50	5	63.8069	0.6603

TABLE V: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [10, 20, 30, 40, 50]$, $m = 5$, bandwidth $n = 3$.

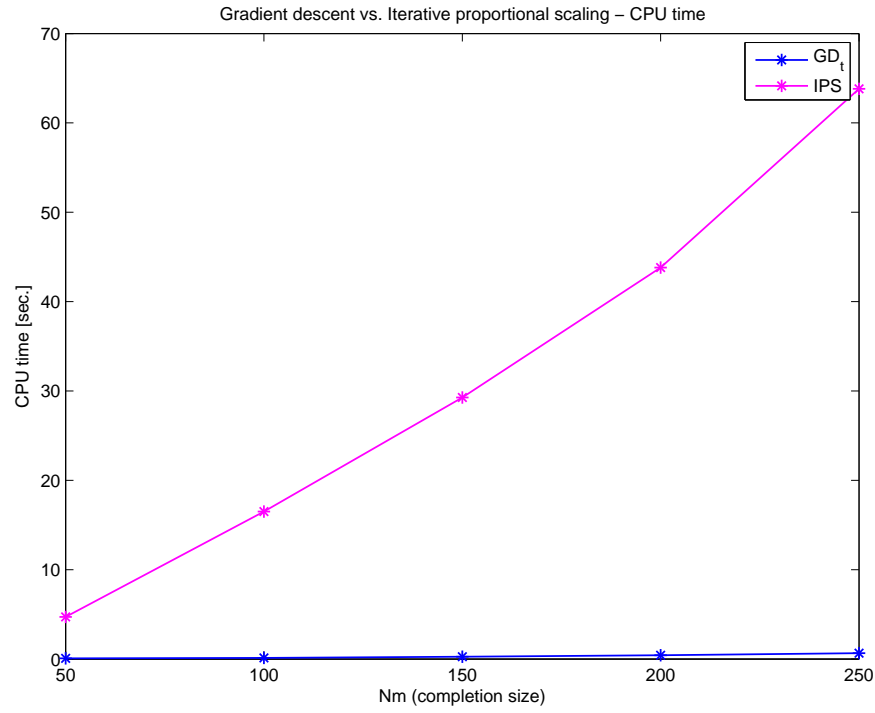


Fig. 7: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [10, 20, 30, 40, 50]$, $m = 5$, bandwidth $n = 3$.

N	m	IPS	GD
10	10	69.7671	0.1516
20	10	307.9596	0.4459
30	10	597.3791	0.8988
40	10	924.6431	1.4798
50	10	1341.0976	2.2052

TABLE VI: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [10, 20, 30, 40, 50]$, $m = 10$, bandwidth $n = 3$.

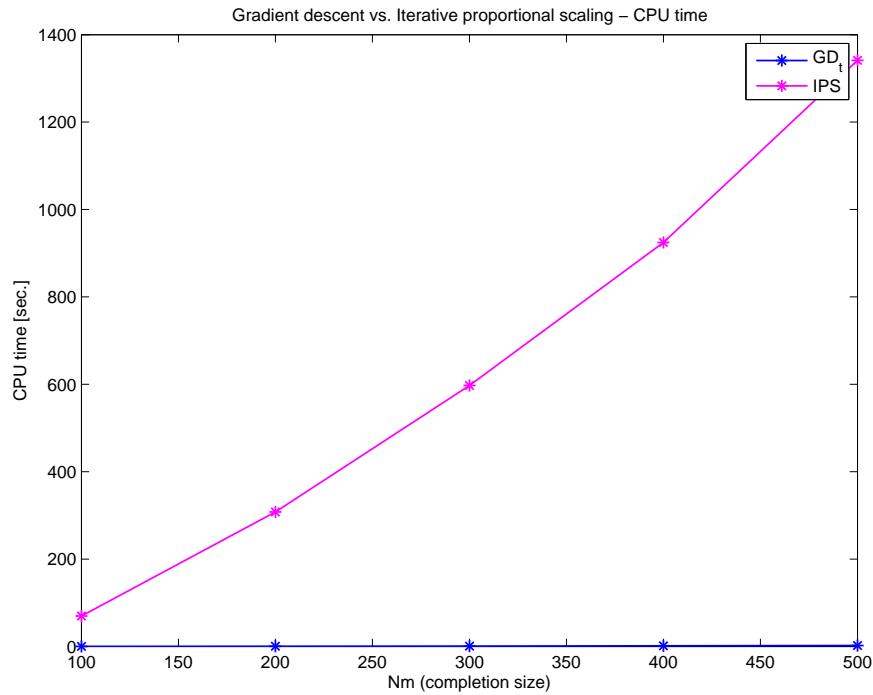


Fig. 8: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [10, 20, 30, 40, 50]$, $m = 10$, bandwidth $n = 3$.

VI. A NECESSARY AND SUFFICIENT CONDITION FOR THE FEASIBILITY OF THE CMEP FOR UNITARY BLOCK-SIZE AND BANDWIDTH ONE

In this section, we provide a necessary and sufficient condition for the feasibility of the CMEP 4 for the special case of unitary block-size and bandwidth. We start by considering a real-valued,

discrete-time, stationary periodic process with circulant covariance matrix. We show that the (stochastic) DFT representation features *uncorrelated* elements. We then write explicitly the covariance lags of such a process. Next, we reformulate the feasibility of the CMEP for such a process in terms of the solvability of a system of linear equations and provide necessary and sufficient conditions for the solvability of this system.

Theorem 6.1: Let $\{y(t)\}$ be a zero mean stationary periodic process of period N taking values in \mathbb{R} . Then

(i) $\{y(t)\}$ can be represented as

$$y(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k e^{jkt \frac{2\pi}{N}}, \quad (34)$$

where

$$c_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t) e^{-jkt \frac{2\pi}{N}} \quad (35)$$

and the c_k 's are uncorrelated random variables.

(ii) the covariance samples of $\{y(t)\}$ are given by

$$\begin{aligned} \sigma_m = \frac{1}{N} \left\{ \mathbb{E} [|c_0|^2] + \sum_{k=1}^{\frac{N}{2}-1} \mathbb{E} [|c_k|^2] 2 \cos \left(km \frac{2\pi}{N} \right) \right. \\ \left. + \mathbb{E} \left[\left| c_{\frac{N}{2}} \right|^2 \right] \cos(m\pi) \right\} \end{aligned} \quad (36)$$

for N even, and

$$\sigma_m = \frac{1}{N} \left\{ \mathbb{E} [|c_0|^2] + \sum_{k=1}^{\frac{N-1}{2}} \mathbb{E} [|c_k|^2] 2 \cos \left(km \frac{2\pi}{N} \right) \right\} \quad (37)$$

for N odd.

Proof: (i) Equations (34), (35) are simply the Discrete-Time Fourier Series equations for $y(t)$. We want to show that the c_k 's are uncorrelated random variables. To this aim, note that (34) can be written in matrix form as

$$\mathbf{y} = V^* \mathbf{c}, \quad (38)$$

where V denotes the Fourier matrix (21) with $m = 1$, and \mathbf{y} and \mathbf{c} are the column vectors obtained by stacking the $y(k)$'s and the c_k 's for $k = 0, \dots, N-1$. Using (38) and recalling that V is unitary, the variance of the vector \mathbf{c} can be written as

$$\mathbb{E} [\mathbf{c}\mathbf{c}^*] = V \Sigma_N V^*, \quad (39)$$

where $\Sigma_N := \mathbb{E} \mathbf{y} \mathbf{y}^\top$. As we recalled in Section IV-A, every circulant matrix can be diagonalized by the Fourier matrix. Since Σ_N is circulant, it follows from (39) that $\mathbb{E} [\mathbf{c} \mathbf{c}^*]$ is diagonal, i.e. the c_k 's are uncorrelated random variables.

(ii) By the uncorrelatedness of the c_k 's, it follows that the covariance lags

$$\sigma_m = \mathbb{E} [y(t+m)y(t)^*] = \mathbb{E} \left[\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k e^{jk(t+m)\frac{2\pi}{N}} \left(\frac{1}{\sqrt{N}} \sum_{\ell=0}^{N-1} c_\ell e^{j\ell t \frac{2\pi}{N}} \right)^* \right],$$

$m = 0, \dots, N-1$, can be written as

$$\sigma_m = \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} [|c_k|^2] e^{jkm \frac{2\pi}{N}}.$$

Taking into account that $c_k = c_{N-k}^*$, we get equations (36) and (37). ■

Theorem 6.1 let us reformulate the feasibility of the CMEP for $m = n = 1$ in terms of the solution of a linear system of equations. This is the content of Corollary 6.1 below.

Corollary 6.1: Let

$$\mathbf{T}_2 = \begin{bmatrix} 1 & \sigma_1 \\ \sigma_1 & 1 \end{bmatrix}$$

be the matrix of the boundary data, where, without loss of generality, we have assumed $\sigma_0 = 1$, and let N be a positive integer, $N > 3$. The data matrix \mathbf{T}_2 admits a positive definite (respectively, a positive semidefinite) circulant completion if and only if the systems

$$\begin{cases} \frac{1}{N} \left(p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k + p_{\frac{N}{2}} \right) = 1 \\ \frac{1}{N} \left(p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k \cos \left(k \frac{2\pi}{N} \right) - p_{\frac{N}{2}} \right) = \sigma_1 \end{cases}, \text{ for } N \text{ even} \quad (40)$$

and

$$\begin{cases} \frac{1}{N} \left(p_0 + \sum_{k=1}^{\frac{N-1}{2}} 2p_k \right) = 1 \\ \frac{1}{N} \left(p_0 + \sum_{k=1}^{\frac{N-1}{2}} 2p_k \cos \left(k \frac{2\pi}{N} \right) \right) = \sigma_1 \end{cases}, \text{ for } N \text{ odd} \quad (41)$$

with $p_k > 0$ (respectively, $p_k \geq 0$), $k = 0, \dots, N-1$, admit solution.

Proof: System (40) (respectively, system (41)) follows simply by writing (36) (respectively, (37)) for $m = 0$ and $m = 1$. ■

We are now ready to state our main result.

Theorem 6.2 (Necessary and sufficient condition - unitary block-size and bandwidth):

The data matrix \mathbf{T}_2 admits a *positive definite* circulant completion if and only if

- $|\sigma_1| < 1$, for N even;

- $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$, for N odd.

Proof: Once again, by Corollary 6.1, it suffices to prove that system (40) for N even ((41) for N odd) with the constraints $p_k > 0$, $k = 0, \dots, N-1$ has solution if and only if $|\sigma_1| < 1$ (respectively, if and only if $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$). For what concern necessity, notice that, for N even, since the p_k 's are $\geq 0 \forall k$

$$p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k + p_{\frac{N}{2}} \geq p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k \cos\left(k\frac{2\pi}{N}\right) - p_{\frac{N}{2}}$$

and indeed, since the p_k 's must be strictly positive, the inequality is strict, i.e. $\sigma_1 < 1$. Similarly,

$$-\left(p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k + p_{\frac{N}{2}}\right) \leq p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2p_k \cos\left(k\frac{2\pi}{N}\right) - p_{\frac{N}{2}},$$

so that $\sigma_1 > -1$ when the p_k 's are strictly positive. Necessity for N odd can be deduced by a similar argument. For what concern sufficiency, the two cases must be distinguish.

Sufficiency for N even: we need to prove that if $|\sigma_1| < 1$, then system (40) with constraints $p_k > 0$, $k = 0, \dots, N-1$ has solution. In fact, setting

$$\begin{cases} p_0 = N(\alpha - \varepsilon) \\ p_{\frac{N}{2}} = N(1 - \alpha - \varepsilon) \end{cases}$$

with $\alpha \in (0, 1)$, $0 < \varepsilon < \min\{\alpha, 1 - \alpha\}$, we get

$$\begin{cases} \frac{1}{N}(p_0 + p_{\frac{N}{2}}) = \alpha - \varepsilon + 1 - \alpha - \varepsilon = 1 - 2\varepsilon \\ \frac{1}{N}(p_0 - p_{\frac{N}{2}}) = \alpha - \varepsilon - 1 + \alpha + \varepsilon = 2\alpha - 1 \end{cases}.$$

Thus if one chooses the remaining p_k in order to satisfy

$$\begin{cases} \frac{1}{N} \sum_{k=1}^{\frac{N}{2}-1} 2p_k = 2\varepsilon \\ \frac{1}{N} \sum_{k=1}^{\frac{N}{2}-1} 2p_k \cos(k\frac{2\pi}{N}) = 0 \end{cases}, \quad (42)$$

then $\sigma_0 = 1$ and $\sigma_1 = 2\alpha - 1$ and the thesis is proved. Since $\sum_{k=1}^{\frac{N}{2}-1} \cos(k\frac{2\pi}{N}) = 0$, it is easy to see that

$$p_k = \frac{N\varepsilon}{(\frac{N}{2} - 1)}, \quad k = 1, \dots, \frac{N}{2} - 1$$

is a solution of (42) and thus

$$\begin{cases} p_0 = N(\alpha - \varepsilon) \\ p_{\frac{N}{2}} = 1 - \alpha - \varepsilon \\ p_k = \frac{N\varepsilon}{(\frac{N}{2}-1)} \quad k = 1, \dots, \frac{N}{2} - 1 \end{cases}$$

with $\alpha = \frac{\sigma_1+1}{2}$, $0 < \varepsilon < \alpha$ solves (40) with constraints $p_k > 0$, $k = 0, \dots, N-1$.

Sufficiency for N odd: the aim is to prove that if $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$, then system (41) with constraints $p_k > 0$, $k = 0, \dots, N-1$ has solution. If we set

$$\begin{cases} p_0 = N(\alpha - \varepsilon) + \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)} \\ p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}} = N\left(\frac{1-\alpha}{2} + \frac{\varepsilon}{2\cos\left(\frac{N-1}{N}\pi\right)}\right) + \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)} \\ p_k = \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)}, \quad k \neq 0, \frac{N-1}{2}, \frac{N+1}{2} \end{cases} \quad (43)$$

with $\alpha \in (0, 1)$, $0 < \varepsilon < \min\left\{-\frac{N\alpha}{1-N-\frac{1}{\cos\left(\frac{N-1}{N}\pi\right)}}, \frac{N}{2} \frac{\alpha-1}{\frac{N-2}{2\cos\left(\frac{N-1}{N}\pi\right)}+1}\right\}$, then the first of (41) is satisfied and the second yields $\sigma_1 = \left[1 - \cos\left(\frac{N-1}{N}\pi\right)\right]\alpha + \cos\left(\frac{N-1}{N}\pi\right)$, i.e. as α varies continuously over the interval $(0, 1)$, σ_1 varies continuously over $(\cos\left(\frac{N-1}{N}\pi\right), 1)$, which concludes the proof. ■

Remark 6.1: Notice that a necessary and sufficient condition for the existence of a positive *semidefinite* circulant completion of the given data can be obtained by simply letting σ_1 to take the values corresponding to the interval endpoints, i.e. the data matrix \mathbf{T}_2 admits a positive *semidefinite* circulant completion if and only if $|\sigma_1| \leq 1$, for N even, and $\cos\left(\frac{N-1}{N}\pi\right) \leq \sigma_1 \leq 1$, for N odd.

Remark 6.2: In their work ([20]) Grone, Johnson, Sa and Wolkowicz show that every partial positive definite matrix (i.e. every matrix such that all the specified principal submatrices are positive definite) has a positive definite completion if and only if the graph associated with the specified entries is chordal ([20, Theorem 7]). However, as we have already noticed, the graph associated with the given entries for the CMEP is not chordal and thus one may wonder if the result of Theorem 6.2 contrasts with the one in [20].

The distinction becomes clear if the necessary and sufficient condition in [20] is rephrased as follows: every partial positive definite matrix has a positive definite completion *for every fixed completion size N* if and only if the graph of the specified entries is chordal. On the other side,

Theorem 6.2 claims that, even for the non-chordal graph associated with the index set \mathcal{I}_b (1) a positive definite completion of the given data may exist provided that the completion size is sufficiently large. So there is no contradiction between the two statements.

VII. EXAMPLE

Example 7.1: Let

$$\mathbf{T}_2 = \begin{bmatrix} 1 & -0.91 \\ -0.91 & 1 \end{bmatrix}.$$

We want to investigate the feasibility of Problem 4 for $N = 7$ and $N = 9$. Since

$$\cos \left\{ \frac{(N-1)}{N} \pi \right\} = \begin{cases} -0.9010 & \text{for } N = 7 \\ -0.9397 & \text{for } N = 9 \end{cases},$$

by Theorem 6.2, we expect that for $N = 7$ the problem is unfeasible while for $N \geq 9$ it is expected to become feasible. For $N = 7$ the set of all positive definite completions is delimited by the intersection of the halfplanes delimited by ([5])

$$\Psi(w^0) = -0.82 + 2x + 2y \tag{44a}$$

$$\Psi(w^1) = \Psi(w^6) = -0.134751 - 0.445042x - 1.80194y \tag{44b}$$

$$\Psi(w^2) = \Psi(w^5) = 1.40499 - 1.80194x + 1.24698y \tag{44c}$$

$$\Psi(w^3) = \Psi(w^4) = 2.63976 + 1.24698x - 0.445042y. \tag{44d}$$

In Figure 9 the intersection Γ of the half-planes identified by (44a) and (44b) is shown, together with the half-plane delimited by (44c). The intersection of these two regions is empty. It follows that the intersection of the four half-planes is also empty, as claimed. On the other hand, if $N = 9$, the eigenvalues are

$$\Psi(w^0) = -0.82 + 2x + 2y + 2z$$

$$\Psi(w^1) = \Psi(w^8) = -0.394201 + 0.347296x - y - 1.87939z$$

$$\Psi(w^2) = \Psi(w^7) = 0.68396 - 1.87939x - y + 1.53209z$$

$$\Psi(w^3) = \Psi(w^6) = 1.91 - x + 2y - z$$

$$\Psi(w^4) = \Psi(w^5) = 2.71024 + 1.53209x - y + 0.347296z$$

and the feasible set is the nonempty region shown in Figure 10.

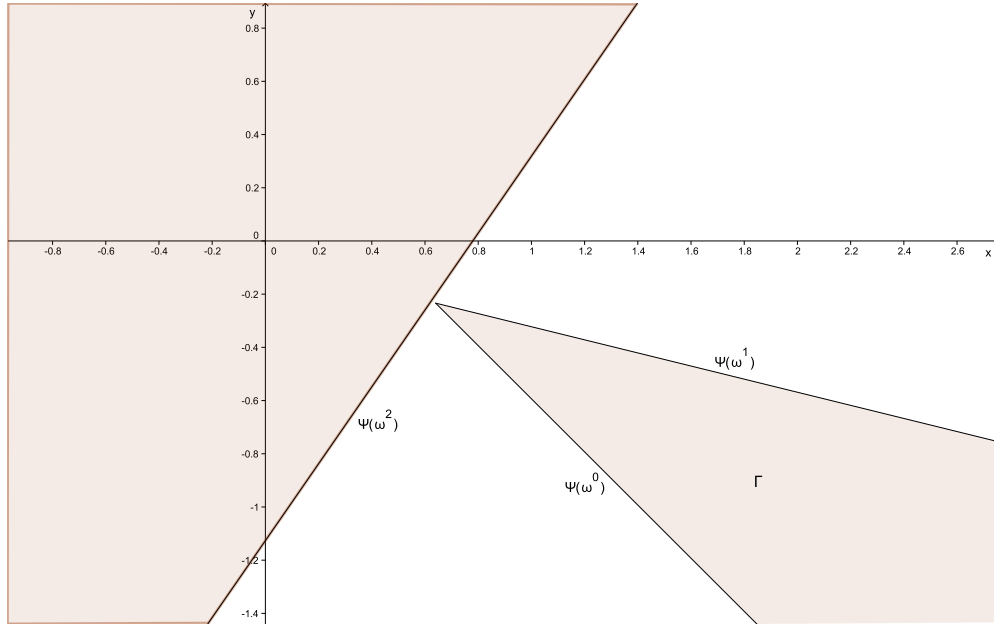


Fig. 9: One half-plane and the intersection of other two of the four half-planes representing the regions where the eigenvalues of $\text{Circ}\{1, -0.91, x, y, y, x, -0.91\}$ are positive.

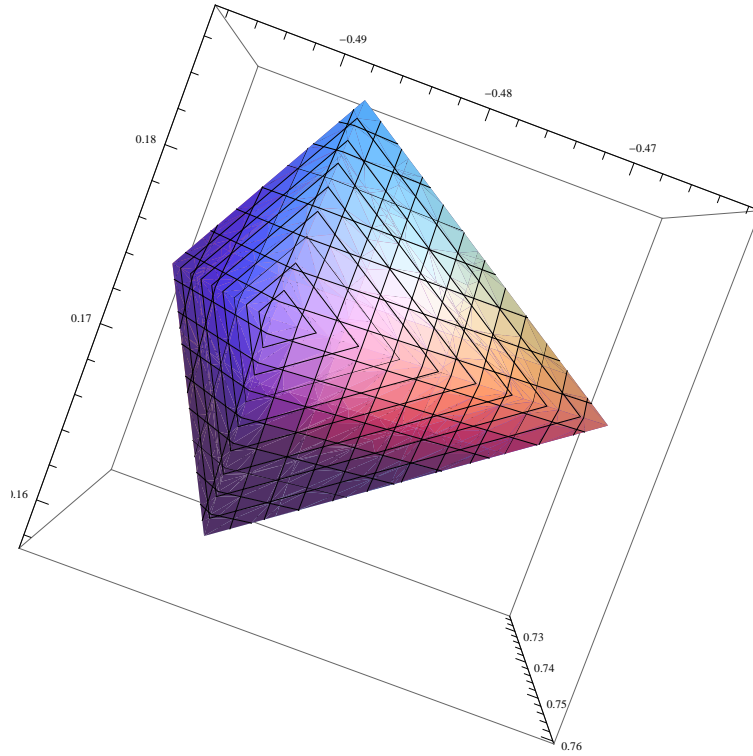


Fig. 10: Feasible region $\{(x, y, z) \mid \Sigma_N \geq 0\}$ for $\Sigma_N = \text{Circ}\{1, -0.91, x, y, z, z, y, x, -0.91\}$.

VIII. CONCLUSIONS

The main contribution of the present paper is an efficient algorithm to solve the maximum entropy band extension problem for block-circulant matrices. This problem has many applications in signal processing since it arises in connection with maximum likelihood estimation of periodic, and in particular quasi-Markov (or reciprocal), processes. Even if matrix completion problems have gained considerable attention in the past (think for example to the covariance extension problem for stationary processes on the integer line, i.e. for Toeplitz matrices), the maximum entropy band extension problem for block-circulant matrices has been addressed for the first time in [4]. The proposed algorithm heavily exploits the circulant structure and relies on the variational analysis brought forth in [4]. Further light is shed also on the feasibility issue for such a problem.

REFERENCES

- [1] W.W. Barrett, C. R. Johnson, and M. Lundquist. Determinantal formulation for matrix completions associated with chordal graphs. *Linear Algebra and Applications*, 121:265–289, 1989.
- [2] S. Boyd and L. Vanderberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] C. Bron and J. Kerbosch. Algorithm 475: finding all cliques of an undirected graph. *Commun. ACM*, 16(9):575–577, 1973.
- [4] F. P. Carli, A. Ferrante, M. Pavon, and G. Picci. A maximum entropy solution of the covariance extension problem for reciprocal processes. *IEEE Transactions on Automatic Control*. to appear, 2011.
- [5] F. P. Carli and T. T. Georgiou. On the covariance completion problem under a circulant structure. *IEEE Transactions on Automatic Control*, 56(4):918 – 922, 2011.
- [6] F. P. Carli and G. Picci. On the factorization approach to band extension of block-circulant matrices. In *Proceedings of the 19th Int. symposium on the mathematical theory of networks and systems (MTNS)*, pages 907–914, Budapest, Hungary, July 2010.
- [7] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.
- [8] J. Dahl, L. Vanderberghe, and V. Roychowdhury. Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods and Software*, 23:501–520, 2008.
- [9] P. Davis. *Circulant Matrices*. John Wiley & Sons, 1979.
- [10] A. Dembo, C. L. Mallows, and L. A. Shepp. Embedding nonnegative definite toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation. *IEEE Transactions on Information Theory*, 35(6):1206 –1212, Nov 1989.
- [11] A.P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.
- [12] H. Dym and I. Gohberg. Extension of band matrices with band inverses. *Linear Algebra and Applications*, 36:1–24, 1981.

- [13] A. Ferrante and B. C. Levy. Canonical form of symplectic matrix pencils. *Linear Algebra and its Applications*, 274(1–3):259–300, April 1998.
- [14] A. Ferrante and M. Pavon. Matrix completion à la Dempster by the principle of parsimony. *IEEE Transactions on Information Theory*, 57(6):3925–3931, 2011.
- [15] M. Fukuda, M. Kojima, K. Murota, and K. Nakata. Exploiting sparsity in semidefinite programming via matrix completion i: general framework. *SIAM Journal in Optimization*, 11:647–674, 2000.
- [16] T. T. Georgiou. Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parametrization. *IEEE Transactions on Information Theory*, 52:1052–1066, 2006.
- [17] I. Gohberg, S. Goldberg, and M. Kaashoek. *Classes of Linear Operators vol II*. Birkhauser, Boston, 1994.
- [18] I. Gohberg, M. A. Kaashoek, and H. J. Woerdeman. The band method for positive and strictly contractive extension problems: An alternative version and new applications. *Integral Equations and Operator Theory*, 12:343–382, 1989.
- [19] M. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York, 1980.
- [20] R. Grone, C.R. Johnson, E. M. Sa, and H. Wolkowicz. Positive Definite Completions of Partial Hermitian Matrices. *Linear Algebra and Its Applications*, 58:109–124, 1984.
- [21] S. J. Haberman. *The Analysis of frequency data*. Univ. Chicago Press, 1974.
- [22] B. Jamison. Reciprocal processes: The stationary gaussian case. *Ann. Math. Stat.*, 41:1624–1630, 1970.
- [23] S. Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968.
- [24] B. C. Levy and A. Ferrante. Characterization of stationary discrete-time gaussian reciprocal processes over a finite interval. *SIAM J. Matrix Analysis Applications*, 24(2):334–355, 2002.
- [25] B. C. Levy, R. Frezza, and A.J. Krener. Modeling and estimation of discrete-time Gaussian reciprocal processes. *IEEE Trans. Automatic Control*, AC-35(9):1013–1023, 1990.
- [26] K. Nakata, K. Fujitsawa, M. Fukuda, M. Kojima, and K. Murota. Exploiting sparsity in semidefinite programming via matrix completion ii: implementation and numerical details. *Mathematical Programming Series B*, 95:303–327, 2003.
- [27] J. A. Sand. *Four papers in Stochastic Realization Theory*. PhD thesis, Dept. of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 1994.
- [28] T. P. Speed and H. T. Kiiveri. Gaussian markov distribution over finite graphs. *The Annals of Statistics*, 14(1):138–150, 1986.
- [29] N. Wermut and E. Scheidt. Fitting a covariance selection model to a matrix. algorithm as105. *Appl. Statist.*, 26:88–92, 1977.
- [30] P. Whittle. On the fitting of multivariate autoregressions and the approximate spectral factorization of a spectral density matrix. *Biometrika*, 50:129–134, 1963.